

## Projet Data science

Une compagnie d'assurance souhaite construire un score d'appétence pour une « garantie accident de la vie », dans un contexte de vente croisée

Les données sont issues de diverses filiales du groupe : autres produits détenus, éléments socio-démographiques, habitudes de vie ou de consommation ainsi que la sinistralité observée.

Les opérations suivantes ont été réalisées sur les données :

- Suppression des catégories rares
- Discrétisation des variables numériques par quantile
- Pas de dimension de temps

L'appétence est désignée par la colonne « target », binaire. Il faut donc prédire les lignes avec une target à 1.

La performance sera mesurée par le lift : % de client appétent identifiés dans les 10% de score le plus élevé.

### A réaliser

#### 1) Prédire le comportement client sur le jeu de Validation (2/3 de la note)

- 3 prédictions minimum sur le jeu de validation
  - Pour chacune de ces prédictions, livrer
    - o Un fichier csv
      - Nom Predict\_**VotreNom\_Modele**.csv (respectez la casse)
      - 2 colonnes (pas 3 .. Vérifiez)
      - Garder le « . » comme séparation décimale
    - o Dans le document accompagnateur indiquer pour chacun de ces modèles
      - Les réglages réalisés
      - Votre estimation d'AUC
    - o Le code
- #### 2) Décrire ce que vous avez compris des drivers de l'appétence client (même si le nom des variables n'est pas explicite) (1/3 de la note) :
- Quelles sont les (combinaisons de) variables les plus importantes
  - Visualisations du comportement associé à ces variables importantes (quel est leur impact sur target ?)

### Le livrable comprend donc :

- Un word (1 page peut suffire, mais sentez-vous libre de vous exprimer)
- Un code
- Autant de csv que de modèles